WHAT IS CLAIMED IS:

1. A method of ascertaining phoneme speech unit boundaries of adjacent speech units in speech data, the method comprising:

receiving training data of speech waveforms with known boundary locations of phoneme speech units contained therein;

processing the speech waveforms to obtain multi-frame acoustic feature pseudo-triphone representations of a plurality of pseudo-triphones in the speech data, each pseudo-triphone comprising a boundary location, a first phoneme speech unit preceding the boundary location and a second phoneme speech unit following the boundary location;

clustering the multi-frame acoustic feature pseudo-triphone representations as a function of acoustic similarity in a plurality of clusters;

training a refining model for each cluster;

receiving a second set of data of speech waveforms with initial boundary locations of adjacent phoneme speech units contained therein;

identifying pseudo-triphones in the second set of data and corresponding refining

models for each of the pseudo-triphones; and

using the refining model for each corresponding pseudo-triphone to locate a boundary location different than the initial boundary.

2. The method of claim 1 wherein clustering comprises maintaining a minimum number of multi-frame acoustic feature pseudo-triphone representations greater than one in each cluster.

3. The method of claim 1 wherein clustering comprises controlling a number of clusters created.

4. The method of claim 1 wherein clustering comprises using a Classification and Regression Tree clustering technique.

5. The method of claim 1 wherein clustering comprises using a Support Vector Machine (SVM) clustering technique.

6. The method of claim 1 wherein clustering comprises using a Neural network (NN) clustering technique.

7. The method of claim 1 wherein clustering comprises using a vector quantization (VQ) clustering technique.

8. The method of claim 1 wherein processing the speech waveforms to obtain multi-frame acoustic feature pseudo-triphone representations comprises forming a multi-dimensional matrix or vector based on a number of frames of speech waveform data adjacent to the known boundary.

9. The method of claim 8 wherein forming a multi-dimensional matrix or vector comprises reducing the number of dimensions.

10. The method of claim 1 wherein training a refining model for each cluster comprises forming a Gaussian Mixture Model to model the most likely locations of a boundary for each cluster.

11. The method of claim 10 wherein forming a Gaussian Mixture Model to model the most likely locations of a boundary for each cluster comprises obtaining only a single Gaussian component.

12. The method of claim 1 wherein training a refining model for each cluster comprises forming a Neural Network model to model the most likely locations of a boundary for each cluster.

13. The method of claim 1 wherein training a refining model for each cluster comprises forming a Hidden Markov Model to model the most likely locations of a boundary for each cluster.

14. The method of claim 1 wherein training a refining model for each cluster comprises forming a Maximum Likelihood Probability model to model the most likely locations of a boundary for each cluster.

15. A computer-readable medium having computer-executable instructions for processing speech data, the computer-readable medium comprising:

an acoustic feature generator adapted to receive training data of speech waveforms with known boundary locations of phoneme speech units contained therein and generate multi-frame acoustic feature pseudo-triphone representations of a plurality of pseudo-triphones in the training data, each pseudo-triphone comprising a boundary location, a first phoneme speech unit preceding the boundary location and a second phoneme speech unit following the boundary location;

a clustering module adapted to receive the multi-frame acoustic feature pseudo-triphone representations of the plurality of pseudo-triphones and cluster the representations based acoustic similarity; and

a refining module generator adapted to operate on each cluster of

representations and generate a statistical model therefor.

16. The computer-readable medium of claim 15 wherein the clustering module comprises Classification and Regression Tree clustering module.

17. The computer-readable medium of claim 16 wherein the clustering module is adapted to maintain a minimum number of multi-frame acoustic feature pseudo-triphone representations greater than one in each cluster.

18. The computer-readable medium of claim 16 wherein the clustering module is adapted to control a number of clusters created.

19. The computer-readable medium of claim 15 wherein the clustering module comprises a Support Vector Machine (SVM) clustering module.

20. The computer-readable medium of claim 19 wherein the clustering module is adapted to maintain a minimum number of multi-frame acoustic feature pseudo-triphone representations greater than one in each cluster.

21. The computer-readable medium of claim 19 wherein the clustering module is adapted to control a number of clusters created.

22. The computer-readable medium of claim 15 wherein the clustering module comprises a Support Vector Machine (SVM) clustering module.

23. The computer-readable medium of claim 22 wherein the clustering module is adapted to maintain a minimum number of multi-frame acoustic feature pseudo-triphone representations greater than one in each cluster.

24. The computer-readable medium of claim 22 wherein the clustering module is adapted to control a number of clusters created.

25. The computer-readable medium of claim 15 wherein the clustering module comprises a vector quantization (VQ) clustering module.

26. The computer-readable medium of claim 25 wherein the clustering module is adapted to maintain a minimum number of multi-frame acoustic feature pseudo-triphone representations greater than one in each cluster.

27. The computer-readable medium of claim 25 wherein the clustering module is adapted to control a number of clusters created.

28. The computer-readable medium of claim 15 wherein acoustic feature generator is adapted to form a multi-dimensional matrix or vector based on a number of frames of speech waveform data adjacent to the known boundary.

29. The computer-readable medium of claim 28 wherein the refining module generator is adapted to form a Gaussian Mixture Model to model the most likely locations of a boundary for each cluster.

30. The computer-readable medium of claim 29 wherein the refining module generator is adapted to form a Gaussian Mixture Model having only a single Gaussian component to model the most likely locations of a boundary for each cluster.

31. The computer-readable medium of claim 15 wherein the refining module generator is adapted to form a Neural Network model to model the most likely locations of a boundary for each cluster.

32. The computer-readable medium of claim 15 wherein the refining module generator is adapted to form a Hidden Markov Model to model the most likely locations of a boundary for each cluster.

33. The computer-readable medium of claim 15 wherein the refining module generator is adapted to form a Maximum Likelihood Probability model to model the most likely locations of a boundary for each cluster.

34. The computer-readable medium of claim 15 and further comprising:

a boundary segmentation module adapted to receive the statistical model for each cluster of representations and a second set of data of speech waveforms with initial boundary locations of adjacent phoneme speech units contained therein and using the statistical models obtain new boundary locations for the adjacent phoneme speech units.